7701-EN-01

UNSUP: AN APPROACH TO UNSUPERVISED CLASSIFICATION OF REMOTE SENSING IMAGERY

Final Technical Report

by

Robert J. Mokken
(December 1998)

United States Army

EUROPEAN RESEARCH OFFICE OF THE U.S. ARMY

London England

CONTRACT NUMBER N68171-95-C-9124

K. P. Piena

TITLE PAGE SF 298

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	7-01-1999	Final report

4. TITLE AND SUBTITLE
Remote Sensing, Statistics and Artificial Neural Networks

5. FUNDING NUMBERS
N-68171-95-C-9124

6. AUTHOR(S)
Prof.dr R.J.Mokken

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CCSOM/ PSCW / Universiteit van Amsterdam
Sarphatistraat 143
1018 GD Amsterdam NL

8. PERFORMING ORGANIZATION REPORT NUMBER
BAA RM 99-1

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

see next page.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Remote Sensing, Statistics, GIS, Neural Networks, Classification methods, Optimization	21
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified		

ABSTRACT

Multispectral remote sensing (RS) images are high-dimensional , their dimension varying from 7 (Landsat TM) to 256 or more for hyperspectral data (AVIRIS). A high spatial resolution leads to huge data volumes for RS datasets. Their interpretation, given the usual lack of sufficient ground knowledge, depends on feature detection , requiring efficient unsupervised classification methods. An approach is sketched using fast *kd*-tree algorithms, with adaptive *k*-NN density estimators, leading to a 4-step unsupervised classification. In the *first* step an adaptive, spatially biased learning sample of spectral values is drawn from an RS image to provide an optimal base for class detection. In the *second* step a density-based cluster analysis detects the class system, for various values of separation and sample coverage. In the *third* step classes are not just defined as labels but also as linear segments on their singular value decompositions. Finally, in the *fourth* step the full image is classified by mapping its pixels to the nearest class as spectral mixtures.
A prototype was developed with Java JDK 1.1 and tested on a Landsat TM image of the Painted Rock reservoir. Performance was quite satisfactory. The resulting image classifications showed good discrimination and class texture.

KEYWORDS

Image processing, feature detection, adaptive sampling , nearest neighbor estimator, multivariatedensity estimation, spatial statistics, remote sensing, multispectral data, geographic information systems, nonparametric modeling, *kd*-algorithms, cluster analysis, multiresolution analysis, wavelets, supervised classification, unsupervised classification,

TABLE OF CONTENTS

LIST OF APPENDICES

UNSUP: AN APPROACH TO UNSUPERVISED CLASSIFICATION OF REMOTE SENSING IMAGERY

1. Introduction

This report concerns the results of a methodological study of statistical classification techniques for remote sensing imagery. This covered the period from 1 September 1995 to 31 October 1998 at the Center for Computer Science in Organization and Management/Applied Logic Laboratory (CCSOM/ALL) of the Faculty of Social Sciences, University of Amsterdam, with participation of the CWI (National Research Institute for Mathematics and Computer Science), Amsterdam, and with the collaboration of the Remote Sensing/GIS Center, U.S. Army Cold Regions Research & Engineering Laboratory (CRREL RS/GISC), Hanover (NH). The research has been sponsored in part by the United States Army through its European Research Office, contract NO. N68171-95-C-9124.

1.1. Background.

The project was based on a proposal for a program of research concerning the application of statistical methods and neural network based techniques to remote sensing imagery, presented at the *International Symposium on Spectral Sensing Research '95* (ISSSR); *Theme: Crisis Support*, 26 November-1 December 1995, Melbourne, Victoria, Australia (Mokken, 1995). In that program it was asserted that recent surveys of the state of the art in the development and use of geographical information systems emphasize the failure of classical statistical methods and theory to contribute effectively to an adequate development of GIS-able spatial analysis (see, for instance, Eurostat, 1994). Conceptually and computationally these can't cope sufficiently with the characteristics of modern remote sensing and other GIS-able data. Openshaw (1994) even speaks of 'statistical hangovers', due to the circumstance that the classical statistical framework is essentially *a-spatial*. At the same time an explosive development in statistics was taking place, providing new, highly computer oriented methods and tools, which are appropriate to tackle just such problems of massive and complex data structures, which thus far mostly evaded analytical solutions (Tanner, 1993). Moreover, a parallel and associated development of computational models emerged in the context of artificial intelligence (AI) and neural computing with equally hopeful promises of useful application in spatial analysis. Recent research has revealed some common statistical underpinnings of these two latter streams, which may be crucial for a further understanding and effective application in spatial analysis. (Cheng and Titterington, 1994; Cherkassky, V. *et al.*, 1994; Ripley, 1993; Wasserman, 1993; Weiss and Kulikowsky, 1991).

Consequently the proposal, as finally awarded for this project, intended to investigate common statistical underpinnings of state-of-the-art neural computing and modern statistical modeling and practice. Its focus was on applications for the processing of Remote Sensing (RS) data in the perspective of Geographic Information Systems (GIS). In particular it concerned pattern recognition and adaptive classification, in relation to feature extraction. It was primarily to deal with such techniques as associative, selforganizing Artificial Neural Networks (ANN's) within the context of unsupervised learning and adaptive search techniques. It thus took two perspectives, statistical and probabilistic design of ANN architectures on the one hand and statistical modeling and tuning of ANN-like structures on the other. The project was meant to lead to the development of, and experimentation with, prototypical applications, supplemented with adequate technical documentation.

1.2. Staff.

Prof. dr. Robert J. Mokken (CCSOM/ALL) was the principal investigator and projectleader for this research and the author of this report. He was assisted by drs. Cees H.M. Van Kemenade (research assistant and PhD candidate computer science at CWI).
Throughout its various stages of research a prototype system of unsupervised classification was developed and implemented by Van Kemenade under the supervision (algorithms) of dr. ir. J. (Han). A. La Poutré (senior researcher Software Engineering at CWI) and of the principal investigator (statistical methods). It is based on data analytic methods and theory studied and applied by Van Kemenade in concert with the principal investigator and using RS imagery data provided for test purposes by CRREL RS/GISC.

1.3. Overview

The central theme, unsupervised classification, will be explained in section 2. Sections 3 sketches the major development stages of the project during the period 1 September 1995 - 31 October 1998, followed by a synopsis of the theory and major elements of the resulting prototype UNSUP in section 4. The report is closed in section 5 with some concluding remarks and recommendations for future research.

2. Research focus: unsupervised classification

Remote sensing data call for high-dimensional pattern recognition techniques based on adequate classification techniques. (Richards, 1993). These can be distinguished in two basic modes: *supervised* and *unsupervised.*

2.1 Supervised classification.

In *supervised classification*, ground cover types or classes are pre-established and defined by external knowledge or judgement. An allocation function or classifier is established by some prior calibration or training procedure based on available special purpose spectral libraries or a limited number of relatively scarce and costly ground samples. Using this allocation function as a classifier all the pixels in the RS image are then classified, *i.e.* assigned to one of the predefined classes.
Various methods can be used in unsupervised classification (Richards, 1993). Statistical methods are usually based on linear discriminant functions and optimal Bayesian classification using maximum likelihood (ML) algorithms. More recently non-linear artificial neural networks have been applied with promising results (Eurostat, 1994; Kanellopoulos *et al*, 1991, 1992).

Usually two stages are distinguished in unsupervised classification. The first stage consists of the training and testing steps, where the classifier is calibrated, the second step consists of the actual classification of pixels in the target RS images.
The first steps define the generalizability of the trained classification procedure to the extensive RS data on which they are to be applied for pattern recognition. Uncertainty and error in the training sample involves the risks of error propagation in the actual data. The external knowledge, usually limited or incomplete, thus distributes its particular innate bias across these data, possibly and likely beyond its unknown domain of generalization. For RS data, where the size of training sets is very small with respect to the size of the full image data to be analyzed, these risks can be very large.
Minimization of those risks will often require conditions which are almost impossible to satisfy for a priori ground sampling.
We shall see in section 5.3 that these risks may be restricted by applying a *supervised* trained assignment of a specific known type of groundcover only to those pixels belonging to previously *unsupervised detected* classes of the image, which are known to correspond tot that groundtype.
Moreover, our partners at CRREL RS/GISC stated that in their experience usually no advance external information will be sufficiently available to use supervised classification methods, for virtually all the RS based GIS applications they meet in practice, such as in crisis management and environmental monitoring.

2.2 Unsupervised classification

For these reasons we concluded that only an approach of *unsupervised classification* should be the focus of our research. Here, no prior external (ground)knowledge is used in the classification method itself. All feature, ground type or class defining information is assumed to be generated by the RS image itself. Here the full set of RS data drives the training step of detection and definition of classes in a RS image, using the spatial and spectral distribution in that particular image.
The technique amounts here to a segmentation of the full image data into relatively homogeneous, separable clusters or classes. Validation of classes and their GIS-oriented interpretation will then be guided by *a posteriori* information (ground samples) and other ancillary external information, obviously a complex task as well.
We didn't find much research concerning unsupervised classification methodology for RS imagery. Usually it seemed to consist mostly of a straightforward application of conventional cluster analysis techniques to RS data sets. In the area of artificial neural network theory self-organizing nets (Kohonen, 1997) offers an alternative approach, which seems closely related to the many techniques of

multidimensional scaling (Cox and Cox, 1994) available for such purposes. These latter predominantly focus on lower-dimensional mappings of points, to which then cluster analysis can be applied in later stages.

However, we sought to develop methods which fitted more closely to the two basic characteristics of the spatial data which we find in RS imagery and GIS utilization: *spatial dependence*, and *spatial heterogeneity*.
Spatial dependence or *spatial autocorrelation,* according to the First Law of Geography (Tobler, 1979, Legendre, 1993), implies strong local association among neighboring observations (pixels), which declines with distance. Any valid spatial analysis should incorporate this feature in its basic premises. Yet most pattern recognition of remotely sensed image data usually proceeds by processing each pixel's information separately and independently over the entire image, thus neglecting this basic feature. Most recent types of spatial analysis attempt to incorporate this local spatial dependence according to two approaches:

1. the *neighborhood* approach, usually in the form of a spatial weight matrix $\boldsymbol{W}$ with elements w_{ij}, where for element (pixel) i, $w_{ij} = 1$ for element j when j is contiguous (a neighbor) of i, and $w_{ij} = 0$ otherwise ($w_{ii} = 0$ usually) (Anselin, 1988).

2. the *distance* approach, where a distance function δ_{ij} defines some distance metric between elements (pixels) i,j in the image space (Cressie, 1991).

Spatial heterogeneity or *non-stationarity* is evidenced by the characteristic non-continuous variation of features across local environments in an image, evidenced by occasionally sharp or disjoint, then erratic, then fuzzy boundaries and contours of objects, areas or regions in images (Dutilleuil and Legendre, 1993).
As a consequence of these two basic characteristics, spatial distributions do not fit the standard distributions of classical inferential parametric statistics and its assumptions of linearity, stationarity and i.i.d. (independently and identically distributed) variates.
We therefore decided to make use of recent developments in nonparametric statistics and multivariate density estimation. (Silverman, 1986; Scott, 1992).

3. Development of project

Two preliminary studies were done leading up to the final stage of this project:

3.1. Wavelet transforms.

In the *first* study we investigated the utility of the recently emerging techniques of *wavelet analysis* and its multiresolution potential for unsupervised remote sensing analysis.
In image and signal processing Fourier transforms and especially Fast Fourier Transforms (FFT's) are predominantly used for data transformation and data reduction in, for instance, spatial frequency enhancement and analysis. (Richards, 1993; Smith, C., Pyden, N., and Cole, P., 1995). Quite recently a new type of transform, the *wavelet transform*, originating in France (the mathematical development of *ondelettes)* in the late eighties (Daubechies, 1992; Meyer, 1990; Koornwinder, 1993), has emerged as an alternative method with many possibilities and is gaining extensive application in many areas, including image processing. Wavelet transforms have a number of attractive advantages over FFT's. The basis functions of Fourier transforms have infinite support and catch the global structure of the signal distribution space (as the sinoids have infinite support), which can lead to undesirable effects in the case of signals which are localized in space, such as in image processing. Wavelets have finite support and are particularly suitable for such locally distributed signals because they are not only localized in frequency but also in space. Their operation can be seen as a form of multiresolution analysis, based on recursive orthogonal matrix transformations, hence computation is relatively easy and fast.
For these reasons we decided to first investigate its applicability for our prior stage of unsupervised structure detection. Hence our first effort to find structures in images were based on the wavelet transform by means of the Haar wavelet. The results are given in Appendix 1 (Van Kemenade, La Poutré and Mokken, 1997).

This approach did certainly help us in getting a better understanding of image processing and insight into the ways wavelet analysis can be applied in spectral analysis. However, we had to conclude that for our specific problem there are a few disadvantages of using Haar wavelets to detect (homogeneous) regions:
- stretched regions, such as rivers and roads, are difficult to track;
- noise and isolated deviant pixels (such as a roof in a forest) can easily result in a homogeneous region being split into many small rectangles, so that more regions are detected than necessary;
- arbitrary shapes can not directly be located accurately. The wavelet transform tends to pinpoint a homogeneous inner region, as the Haar wavelets have a square support, and therefore only locate a large square within the region. These drawbacks are related to the fixed choice for the shape of the support, being rectangles of different sizes at fixed locations.
However, different forms and applications of wavelet transforms to other problems of remote sensing analysis may well prove to be promising for the future. Given the main purpose of our present research we decided not to pursue wavelet analysis further for this project, but to concentrate directly on more efficient unsupervised spectral feature detection.

3.2. Bayesian cluster analysis.

A survey of unsupervised Bayesian oriented classification was then made for our *second* study. Beyond standard cluster analysis we found this area of unsupervised classification in image processing to be scarcely covered by other current state-of-the-art-techniques. In classical cluster analysis the number of clusters is fixed or postulated beforehand and then the cluster structure, best fitting that number, is found. In unsupervised searching situations, however, there is no reason to expect or impose *a priori* a certain number of clusters, so we sought to find and develop a method where the ultimate number of classes or clusters is determined simultaneously with the cluster structure.
In doing so we obtained and studied a recent public domain state-of-the-art method and program of unsupervised Bayesian classification (AutoClass: Cheeseman and Stutz, 1996; Stutz *et al*, 1996). It was based on Bayesian mixtures of multivariate distributions, covering continuous (multivariate normal *i.e.* Gaussian priors) as well as discrete (Dirichlet priors) distributions. Classes were estimated by a maximum likelihood (ML) optimization of posterior distributions. This approach had sufficient scope to be useful for our purposes.
We also developed a procedure of biased pixel-sampling, based on locally homogeneous spatial neighborhoods. We needed that for an efficient initial unsupervised classification, using Autoclass, in order to find the basic spectral classes of an image, as determined by its specific local homogeneity structure. Work was then done to map the resulting pixel data backwards into the original image by means of Autoclass. The resulting classifications, based on varied simulated images, looked quite sensible at first sight.

3.3. Nonparametric remodeling.

However, our results with AutoClass, though encouraging as to general strategy, left room for some doubts concerning its efficacy in the specific area of RS multispectral imagery. In particular the Gaussian elements in the model do not always match the rather fragmented and discontinuous nature of such data. Substitution into the Autoclass package as such of nonparametric, *e.g.* density estimation based methods for such Gaussian elements or modules was far from feasible. The re-engineering of the input formats for our RS data sets, necessary to ensure adequate data portability, proved an excellent context to do some re-engineering of the unsupervised classification procedures as well. So we undertook the experimental construction of an entirely nonparametric stepwise, but integrated clustering system, based on the approach of density estimation techniques and designed a/o to circumvent such problems as those due to the fact that the Gaussian model tends to result in the detection of too many classes, and the instability and sensitivity of principal component based data reduction for outlying pattern irregularities.
We started studying density estimation oriented theory (Silverman, 1986; Scott, 1992) and applied that in the development of our final prototype *UNSUP*.

4. UNSUP: a system for unsupervised classification of RS imagery.

The final stage of this project concerned the integration of the various modules into *UNSUP*, a prototype system for unsupervised classification of multispectral images in remote sensing. We shall confine us here to a summary of its main elements. A more elaborate account of the underlying theory

and method is given in Appendix 2, (Van Kemenade, La Poutré and Mokken, 1998a), and a full description of the system UNSUP and its operation is given in Appendix 3, (Van Kemenade, La Poutré and Mokken, 1998b).

4.1. Challenge and compromise: the two curses

Multispectral RS images are high-dimensional data, their dimension m varying from a typical value of $m = 7$ for the seven wavelengths of Landsat Thematic Mapper images up to the hyperspectral data from imaging spectrometers, such as AVIRIS, where the number of bands $m = 256$ or more.
This, together with the increasingly high spatial resolution of these images, produces the huge data volume, which characterizes the data sets of RS imagery . As a consequence statisticians and other data analysts who seek to analyze such sets, are challenged by what have been called two 'curses'. These are the curse of dimensionality (Bellman, 1961) and the curse of optimality (Scott, 1992).

The *curse of dimensionality* refers to the sparse and eccentric distribution of points in high-dimensional space (*i.e.* $m > 4$), and the resulting difficulty to define and detect structure in terms of usual spatial intuition (Scott, 1992). We can see that, if we look at what happens to a classic indicator of the neighborhoods of a point in m-space, the ball. The volume V_m of the m-dimensional unit hypersphere reaches a maximum for $m = 5$ and then dwindles quickly to zero as m increases. Consider any hypercube in m-space and the hyperball inscribed into it. Then the ratio of the volume of that ball to that of its cube approaches zero for increasing m; that is, virtually all the volume of the cube is in its corners, relative to the ball. A similar situation is seen when we consider two concentrically nested m-balls (same center, radii r and $r - \varepsilon$, respectively). Then for large m almost all volume of the larger ball is in the shell between them, approximately a $(m - 1)$ - dimensional surface.
This tendency of mass to move away to the margins of m-space for large m has its consequences for the location of probability mass in multivariate distributions of points in high-dimensional space. For instance, for $m >> 5$ most of the probability mass of multivariate normal ('Gaussian') distributions is in their tails, *e.g.* for $m = 10$, 99% of total probability mass is in the 'tails' (Silverman, 1986).
As a consequence, when developing and interpreting data analytic methods to define and search structure in such high-dimensional contexts, we had to account for these eccentric effects in m-space.

The *curse of optimality* refers to the tendency to use classical optimal or 'efficient' estimation and other computation methods in practical situations where more general or less 'efficient' methods in fact perform at least as well. For instance, in multivariate density estimation there are numerous differing kernel estimation techniques, many of which, though theoretically superior, are computationally too intensive to match the practical requirements of analyzing the megasets of multivariate pixels which characterize RS image data in reasonable time, if not online. We therefore had to compromise between the theoretical requirements of statistical efficiency and accuracy on the one hand and those of adequately fast processing performance on the other. Accordingly we needed fast density estimation methods leading to reasonable approximations of actual class densities in images. Actually only rough estimates will be sufficient, as the main emphasis will be on the location of density modes, instead of accurate density contours. Hence, the application of corresponding, fast algorithms is of prime importance here. Many of these are based on *kd*-trees, advanced data structures, which are multidimensional extensions of the binary tree, where different levels of the tree uses a different dimension for discrimination (Bentley, 1975). For our purposes we made use of optimized *kd*-trees (Friedman, Bentley and Finkel, 1977).

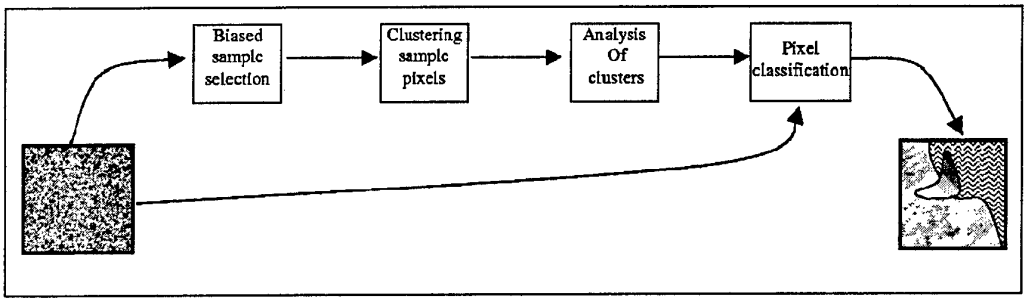
To conclude: against this background of methodological challenges and the necessity to compromise our principal research efforts concerned:
- the development of an effective biased sampling procedure, taking into account the local geospatial distribution of pixels in an RS image;
- adapting and using ideas from statistical multivariate density analysis throughout the system, using the k^{th} nearest neigbor density estimates (k-NN; Loftsgaarden and Quesenberry, 1965; Dasarathy, 1991) specifically as a means for adaptive spatial sampling and unsupervised class detection in high-dimensional space;
- adapting and developing adequate, fast search algorithms to ensure sufficient performance;
- cross-platform portability of the system with respect to Unix (Sun/Solaris) and Wintel environments, in order to ensure testing and use both in the field (laptops) and on PC's, as in high end workstation environments. For that reason development was in Java.

4.2 Basic outline: four steps.

We suppose the image to be *m*-spectral, *m* typically corresponding to 7-band multispectral RS image data. Hyperspectral 256 band images may have to be preprocessed to achieve most informative selections or reductions to 5-10 band format. The corresponding classified image mapping will be in terms of the 3-band RGB representation.

The method is based on four steps as shown in Figure 1:
1. adaptive selection of a *biased spectral learning sample* of pixels from (part of) the image for unsupervised detection of classes in the *m*-dimensional spectral space of the image;
2. *unsupervised detection* of spectral classes as *clusters* in that spectral sample;
3. within each of the established classes *analysis* and characterization of its spectral coordinates and distribution in the spectral sample space.
4. mapping the classes to the entire image by *classification of its pixels* in terms of the established system of classes.

Figure 1: Schematic representation of UNSUP



4.2.1. Step 1. Spatially adaptive learning sample for spectral classification.

Instead of analyzing all pixels in the image, a sufficiently large sample of pixel spectra is chosen for various reasons.
In the first place analysis of just an adequately clustered spectral sample from the original image can ensure sufficient speed in performing classifications on large images.
Moreover one would like to select a set of pixel spectral values (spectral vector $\vec{s}_m$) from an image which are most informative concerning the distribution of features and class structure of that image and hence useful for the unsupervised class detection or learning stage. Not all pixels in the image will do for that purpose, because of noise, or due to varying abundance or scarcity of objects and features in it. Just a random sample from the image would select more pixels from dominant ground covers, *e.g.* sand in a desert image, than necessary for the learning stage of class detection and would likely miss the small trail of a brook in that desert.
This is why the clustering and selectivity of the sampling is adaptively biased toward selection of pixels containing one type of groundcover ('homogeneous' or 'pure' pixels), with a low degree of noise. This is achieved by:
- ensuring local geospatial representativeness and coverage over the image, by stratified grid-sampling from (a rectangular part of) the image;
- selecting pixels which are spatially (window space in image) homogenous in the form of small locally adjacent clusters (patches) and then selecting within these patches pixels which are spectrally homogeneous (in the spectral sample *m*-space).

From the four sampling options which were studied, a method based on *local/global density ratio maximizing*, came out as the best for our purposes. Given a projected sample size of *N*, the stratification is performed by partitioning the image by a grid in *N* square or rectangular parts. Within each cell

(stratum) of the grid a small patch of size $l \times l$ is chosen randomly. For that patch iteratively a pixel is selected with the largest value of the local/global density ratio

$$R^{(l,g)}(\vec{s}_m) = \frac{\hat{f}^{(l)}(\vec{s}_m)}{\hat{f}^{(g)}(\vec{s}_m)},$$

as follows.
Within each patch a pixel is randomly selected as a starting point and the pixel with highest density in its k_l-NN neighborhood ('median' or modal point) found. This is repeated for that point until no new points are found within the patch. This defines a modal local density $\hat{f}^{(l)}(\vec{s}_m)$ for that patch. The global spectral density distribution $f^{(g)}(\vec{s})$ is estimated on a simple random sample from the total image and the corresponding global modal density $\hat{f}^{(g)}(\vec{s}_m)$ analogously determined by a search of the k_g-NN neighborhood of $\vec{s}_m$. After some runs with different starting points the spectral value of the pixel with the highest local/global density ratio *R*, typically in the range 10^2 to 10^4, is chosen as the spectral datapoint representing this patch in the learning sample.
In the example of an image of a desert it will be clear that for pixels with sandy neighborhoods the value of *R* will tend to be low, though obviously, sufficient pixels of the potential class 'sand' will be selected in the learning sample, due to the dominance of sand in the image. However, in those squares of the grid covering part of the lonely brook, the value of *R* for pixels covering that brook will tend to be high, so that there is a good chance that the selection runs end up in such a pixel.
Consequently, this sampling method is designed to correct for dominant classes, in order to ensure sufficient representation in the learning sample of small, relatively scarce spectral classes, such as those corresponding with littered small objects, *e.g.* houses, or thin structures, *e.g.* roads, occupying something like 0.1% and 0.3% of the total image, respectively.

4.2.2. Step 2. Spectral cluster analysis: unsupervised class detection on the learning sample.

Our method of adaptive biased sampling results in a sample of *N* multispectral datapoints, consisting of *m*-dimensional spectral vectors $\mathbf{x}_i; i: 1,\ldots, N$. This sample was designed to represent the spectral distribution over its pixels, which is characteristic for the features and class structure in the image. Our method of unsupervised class detection seeks to 'learn', that is to retrieve, that class structure from this *N*-sample. Spectral classes are seen as relatively dense clusters or point clouds in the spectral *m*-space spanned by the *N*-sample. To find and retrieve these classes we use a method of hierarchical cluster analysis based on an adaptive multivariate density estimator, the multivariate *k*-NN estimator. This estimates the density of the *k*-NN neighborhood for each sample point, as given by the *m*-dimensional ball with that datapoint as its center and the (Euclidean) distance to its k^{th} nearest neighbor as its radius. The nearest neighbor estimator is known to perform better than other, fixed kernel estimators for high-dimensional problems ($m \geq 5$: Scott, 1992, 190) and is known to perform quite satisfactory in high-dimensional analysis, such as cluster analysis. The *k*-NN density estimator for a sample point $\mathbf{x}$ is given by:

$$\hat{f}(\mathbf{x}) = \frac{k}{N d_k^m(\mathbf{x}) V_m};$$

where $d_k(\mathbf{x})$ is k^{th}-NN distance of $\mathbf{x}$ and V_m is the volume of the unit *m*-dimensional ball.
Class detection is then performed by means of a variant of hierarchical cluster analysis (Kaufman and Rousseeuw, 1990). First the sample points are ordered according to decreasing *k*-NN density. Eligible points have densities exceeding a given minimum density threshold. In high-dimensional space the 'curse of dimensionality' will hit us with very sparse densities, together with high values of *k*-NN distances.
One criterion for minimum *k*-NN point density is to define the *k-NN* neighbourhood of a sample point $\mathbf{x}^{(s)} \in R^m$ as a confidence set around $\mathbf{x}^{(s)}$, with confidence level $\overline{\alpha}_k$ (*e.g.* $\overline{\alpha}_k = .95$) and containing *k* randomly selected points $\mathbf{x}_j$ which are multivariate normal ('Gaussian') distributed around that sample

point $\mathbf{x}^{(s)}$, with common small error variance $\varepsilon^2 \mathbf{I}_{m\times m}$ and common mean: $\mathbf{x}^{(s)}$. The corresponding maximum radius criterion is then given by:

$$d_k(\mathbf{x}) \le r^*_{m,k} = \varepsilon\sqrt{\chi^{2-}_m\left(\bar{\alpha}_k^{1/k}\right)}\ ;$$

where χ^{2-}_m denotes the inverse chi-square distribution with m degrees of freedom. The corresponding minimum density criterion (not weighted for sample size N) is:

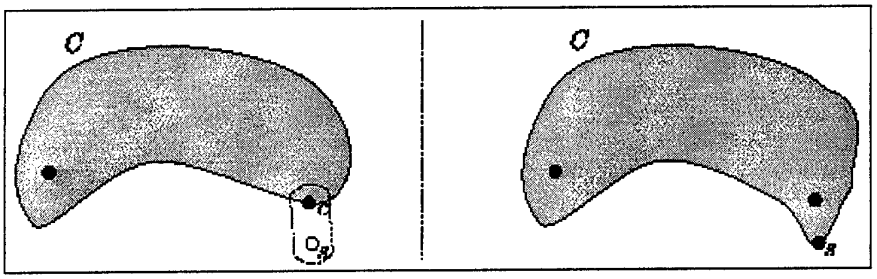
$$k\text{-}NN \text{ density} \ \ge \text{MinDensity} = \frac{k}{\left(r^*_{m,k}\right)^m V_m}\ .$$

An example for the one-dimensional case, compared with the 6-band spectral (Landsat Thematic Mapper) data we used in our experiments: if we work with $k = 10$ and require a small error variance per band of 1% of the band range (0, 255), then we find for the one-dimensional case a maximum k-NN radius of 7.17 and a corresponding unweighted minimum density of 0.70. For the 6-dimensional case these figures are 11.01 and 1.09E-06, respectively. Note the enormous decrease in volume density of the k-NN neigborhood.

A strict, or even 'realistic' value of required minimum density, however, can result in a relatively large fraction of the learning sampling not being allocated to a class. To control that we used a different dominating threshold: the fraction of points allocated to some class. It sets a lower bound on the fraction of sample points to be allocated to one of the eligible clusters . The minimum density criterion is then adapted so that the total fraction of the sample allocated to some cluster at least equals that threshold. Obviously its operation possibly implies considerable leniency concerning the minimum density criteria that follow from its application. It was designed in order to cope better with the 'curse of high dimensionality' and consequently is expected to perform well for higher dimensions (*i.e.* $m \ge 5$).

Sample points s will be merged with a class C in its proximity, when the union of its k-NN neighborhood with that of the nearest point c in that class is sufficiently dense. (See Figure 2).

Figure 2. Merging a point with a cluster



This joint neighborhood of s and c is defined by the *cylindrical envelope* of their k-NN neighborhoods, which is a cylinder with half-spherical sides.
With the margin density for two clusters, c_k and c_l, (*i.e.* the density V of their cylindrical envelope), a criterion for their separability is defined, their *separation*, given by:

$$\frac{V}{\min\{p_k, p_l\}}\ ;$$

where p_i denotes the maximum density in cluster c_i. Clusters are separated when their separation *coefficient* is below a chosen threshold value, *e.g.* 0.50. Otherwise they are merged into one class.

Manipulation of the separation parameter gives the means to vary the number of resulting classes and hence the level of detail in the classified image. In Table 1 we give for various values for the separation threshold the resulting number of classes detected, as found in our experiments with a six band part of a Landsat Thematic Mapper image of the Painted Rock reservoir.

As a result of *Step 2* the detected clusters or classes are found as a partition of the spectral learning N-sample into |C| classes. Each class consists of at least 2 points, because singleton sample points are not considered as a class. However, the pixels in the image, from which these singleton spectral sample points originate, can and will be allocated to some class in the following step of classification.

Table 1. Painted Rock: number of classes for varying separation

	Separation					
	0.125	0.25	0.375	0.50	0.625	0.75
Number of classes	6	11	15	18	24	38

The size of the class is determined by the number of sample points per class. Classes are ordered according to decreasing size, which ordering serves for an assignment of a false color label by descending a RGB color table.
This determines the set of classes detected for the RS image. The next step serves to give an additional characteristic of the location and distribution of pixels (spectral values) within each class.

4.2.3. Step 3. Distribution analysis: linear class segments and blending.

In conventional cluster analysis, such as ISODATA (Ball, 1965; Hall and Ball, 1965; Hall and Khanna, 1977), the unsupervised classification module in the ERDAS-IMAGINE package (Smith, Pyden and Cole, 1995), classes are just defined in terms of labels and located in terms of their centroid.
In UNSUP we wanted to go beyond that, by modelling classes as linear mixtures of spectral values, in order to allow for additional detail within a class. Hence classes are mapped as linear segments, where points within classes are projected as mixtures of the corresponding spectral endpoints.
We did so by using a variant of *projection pursuit* (Friedman and Tukey, 1974; Huber, 1985; Jones and Sibson, 1987), which usually is based on searching optimal projections on subspaces of the full sample space in an attempt to combine dimensional reduction with cluster detection.
For UNSUP we decided to do first the class detection and then characterize each pixel in a class by the first principal component (1st PC) in its Singular Value Decomposition (SVD: Joliffe, 1986). The class is then mapped on a line segment directed by this eigenvector. Hence its linear segment and endpoints are defined by the first m-component eigenvector of the corresponding class, as given by the SVD. Endpoints are chosen symmetric around the centroid of the cluster, so that the length of the cluster segment is equal to 2c times the standard deviation $(2c\sqrt{\lambda_1})$ along the 1st principal component, where λ_1 is its eigenvalue and c is a given constant (*e.g.* $c = 1$, or 1.5).
The corresponding 1st PC-segments can be used:

- to serve as lines of orientation for the classification of pixels in the next step of image classification;
- the projections of the pixels of a class on its first component represent these pixels as mixture coordinates with respect to their segment endpoints. We shall call this *blending*.
- to analyze the distribution of the projections of pixels along these linear class segments.

Testing for unimodality. For a 'pure' class the distribution of projections of its pixels typically should be unimodal (binomial, or continuous mixture of binomials). Multimodality of that distribution could indicate a 'mixed' class. That is why we implemented experimentally testing for multimodality to detect 'impure' mixture classes. This is done with the familiar nonparametric Kolmogorov-Smirnov (K-S)-test, which is based on the maximum distance between the observed cumulative empirical sample distribution and the cumulative sample distribution to be expected under the hypothesis of a distribution with certain modality.
Thus the empirical distribution of pixels along the linear segment of a class is tested against three theoretical distributions: one *unimodal continuous* (the Gaussian or Normal distribution) and two possibly *multimodal* ones (4 point polygon and 8 point polygon approximations).

4.2.4. Step 4. Classification of image: allocation of pixels to spectral classes.

The previous steps led to detection and definition of the typical class structure of the image at the chosen levels of the main parameters: fraction of points classified and separation of classes.
The last and fourth step consists of assigning each pixel in the image to the best fitting class. As the class structure, together with the linear segments and spectral distribution within classes, is contained in the learning sample, we might consider this last step as one of *supervised* pixel classification, for which traditional optimal statistical techniques as linear discriminant analysis are known. These, however, would involve additional training of a discriminant function on the classified learning sample, which would further impede the performance of the total classification process. We therefore used a more direct pixel allocation approach, involving two alternative classification methods.

The *first* method is a nonparametric *nearest neigbor* (*1*-NN) classifier. Each pixel is classified by finding the (spectrally) nearest sample point in the learning sample. The class of the nearest point is assigned to that pixel. Although a *kd*-tree is used to find the nearest neighbor, the classification time per pixel is of order *log N*, where *N* is the size of the spectral learning sample. Hence this is the *'slow classification'* of the two methods.

The *second* method can be seen as a form of *projection pursuit* (PP) classifier, using the 1st PC based on the 1st eigenvector in the SVD of the classes, which we designated above as *blending*.
For each pixel in the image its distance to each class (*i.e.* line segment) is determined in terms of (co-ordinates) of its projection (1st PC) on and the projector (*i.e.* perpendicular distance) to each class's 1st eigenvector and the pixel is then allocated to the nearest class.
Classification time per pixel is of the order of $|C|$, the number of classes. Hence this is the *'fast classification'* of the two methods.

4.3. Three compound classification options.

For the application of UNSUP we have the two options of the 'slow' *1*-NN and the 'fast' PP classifiers. For the *1*-NN classifier we have the option to use the *blending* option in the final image mapping. When the PP classifier is used blending is part and parcel of the mapping.

Consequently in UNSUP we have three compound options for image classification:

option	*classifier*	*blending*
(1)	*1*-NN (slow)	no
(2)	*1*-NN (slow)	yes
(3)	PP (fast)	yes

In *option (1)* image classification is done by the 'slow classification', the direct first nearest neighbor method, with no subsequent blending of pixel class membership. As a consequence the only class information per pixel consists in the pixel labeling, *i.e.* color mapping. The results of this option are comparable to those of traditional cluster analysis, such as ISODATA in the ERDAS system (Smith, Pyden and Cole; 1995).
This option is probably most appropriate in the initial stages of image classification, where the primary validation of the classes is the basic issue, to be performed by investigating the class reproducing properties of various UNSUP.SET steering parameters, in connection with (partial) external knowledge based on other sources or ground investigation.

The two other strategies, *option (2)* and *option (3)* only differ in their method of pixel classification, *slow* or *fast*, respectively. In both cases *blending* provides mixture detail within classes. By blending, any pixel in the image, say with original spectral vector $\boldsymbol{x}$, is represented by its projection $\boldsymbol{y}$ as a mixture of the endpoints of the linear 1st PC segment of its class. These endpoints are given by the sum of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, where $\boldsymbol{a}$ indicates the first point of the 1st PC segment, $\boldsymbol{b}$ the direction vector parallel to the 1st eigenvector of the class cloud in the sample and their sum the second point of the segment. Given a constant $c > 0$, the length of the segment, $|\boldsymbol{b}|$, corresponds to $2c$ times the standard deviation along the 1st PC. Its location is chosen so that $\boldsymbol{a} + .5\boldsymbol{b}$ indicates the centroid of the class cloud in the sample. Pixels, with projections $\boldsymbol{y}$ in between $\boldsymbol{a}$ and $\boldsymbol{b}$ are represented as a mixture $\boldsymbol{a} + k\boldsymbol{b}$, with

$k \in [0,1]$. Pixels with projections outside that range are mapped on the nearest endpoint, so k is then set to 0 or 1. The length of the projector $|x - y|$ is an indicator of the distance of the pixel to its class. In the image mapping within classes, blending is made to correspond with the luminance value of class color. The fraction $k \in [0,1]$ for the pixel is represented in terms of a luminance value ($\in [0,1]$) for the color representing the class label; for $k = 0$ (initial endpoint $\boldsymbol{a}$) the lowest (darkest) value and for $k = 1$ (the second endpoint $\boldsymbol{a} + \boldsymbol{b}$) the lightest.

The final classified image is stored in the form of an RGB image in GIG-format.

5. Concluding remarks and recommendations.

We conclude this report with some comments concerning the implementation and testing of UNSUP, followed by suggestions for further research.

5.1. Implementation and performance.

UNSUP was implemented in Java. All code is compatible with JDK, version 1.0.2., although most of it was compiled using JDK 1.1. For an operational introduction refer to Appendix 3 (Van Kemenade, La Poutré and Mokken, 1998b).
To give a low end indication of the performance of the prototype UNSUP system, as developed in Java, we ran it on part of a Painted Rock Landsat Thematic Mapper image, consisting of 1000 x 1600 pixels for 6 band spectral values. (We omitted Band 6, referring to Richter, 1993). We used a sample of 4000 pixels. The window size for density estimation, *i.e.* neighborhood size was $k = 10$. Separation was set at 0.50.
This was run on a laptop PC, Pentium MMX, 233 MHz, 64 MB RAM, Window 98, including Microsoft Internet Explorer 4.01 and its associated (command line) Java runtime engine.

Approximate times, as observed, were :

Compound option:	*(1)*	*(2)*	*(3)*
Up to completion sample	1 min	1 min	1 min
Class detection and image classification	12 min	15 min	5 min
Total	13 min	16 min	6 min

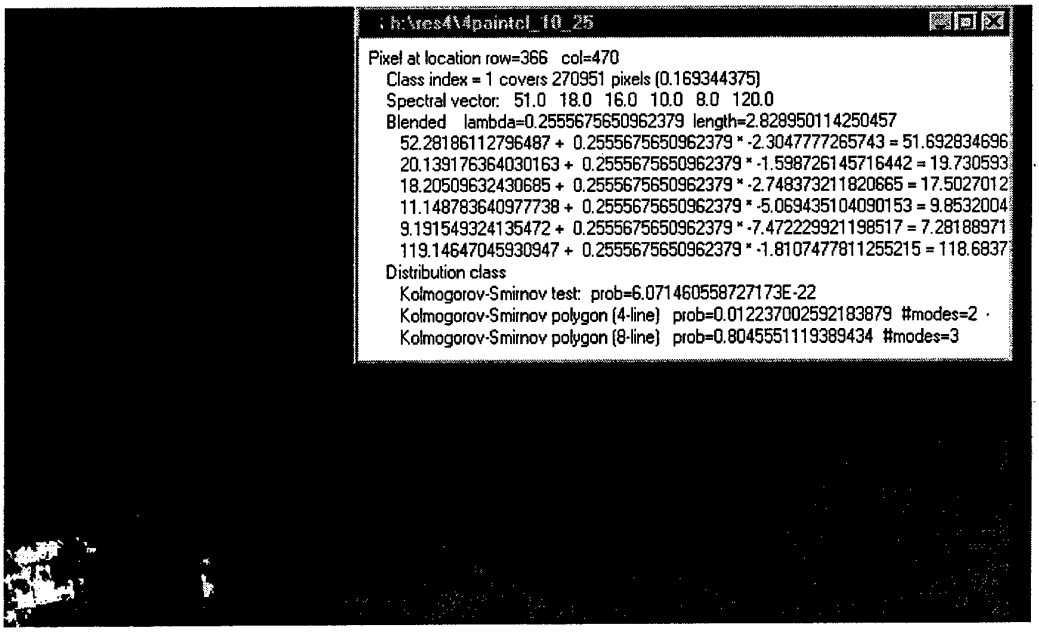
5.2. Experimental testing of UNSUP.

We did an extensive amount of testing with the prototype of UNSUP, with the above mentioned RS image of the Painted Rock reservoir, using a range of parameter values (window sizes and separation values). We shall not treat the results here, as these will be distributed separately on a cartridge, accompanying the execute and source files of prototype UNSUP, version 0.5. An idea of the output is given in Figure 3.

After conclusion of the analysis with classified image and report writing and saving, an option is given for closer image inspection. The classified image is presented, as in Figure 3 (in this text just a copied detail of the actual image), where one can dynamically select a pixel by pointing and clicking with the cursor. A window then gives a list of basic values, characterizing the classification values for that pixel. The following data are given:
- pixel coordinates in image;
- index of allocated class, with absolute and relative size (number of pixels; proportion of image);
- its original (6-band) spectral values;
The then following values are given only for the case of *blending* (options (2) and (3)):
- *blended lambda* gives the mixture coefficient for this pixel with respect to the line segment along the 1st PC of its class, as a fraction of its length between the two endpoints of that segment; followed by the value of the *length* of its projector.
- the following six lines, giving the *mixture coordinates*, describe the six band values of this pixel as a mixture of the endpoint.
- the last three lines report the results of the three modality KS-tests.

During a workshop at CRREL RS/GISC, Hanover, NH, 5-6 October 1998, these tests and results, together with a hands-on demonstration, were discussed with the expert GIS analysts there, with gratifying conclusions.

Figure 3. Detail of classified image of Painted Rock; option (3): PP.



5.3. Further perspectives

UNSUP is a prototype based on a methodological study. To prepare and upgrade it for further use in actual applications, elaborate testing, validation and evaluation by experienced and knowledgeable GIS analysts will be necessary, together with modifications and additions in the area of data management and user interfaces, in order to use it in the context of a data analytic GIS environment.
For instance, features to assign colors to classes at the user's discretion will be necessary to facilitate class comparison across different runs with UNSUP on the same image (*e.g.* with different levels of separation). Adequate database storage of class blending coordinates at the pixel level will be needed to enable further statistical analysis and exploration of class features in an image, using available statistical modules from other sources. More generally, the applicability of UNSUP could benefit from an embedding of the system as a module in a state-of-the-art GIS processing environment, based on UNIX or Windows NT.
Such developments are beyond the methodological scope of our research, which should focus on further methodological improvements and innovation.

Some improvements come to mind after our recent testing experience. For instance, for the definition of the linear class segments there should be different parameter settings, instead of just one for the constant c, which measures distances in their standard deviation: one for pixel classification, and one for blending. Why?
- To ensure robustness of the *class allocation of pixels* it might be better to allocate pixels with respect to the central part of the class point cloud in the sample. Hence the endpoints should be based on the values $c = 1.0$, or 0.68 to cover the central 2/3's or 50% of the point cloud's distribution (Gaussian case).
- On the other hand for mapping by *blending* the pixels within their class, the values of $c = 1.5$ or 2 are preferable, the endpoints then covering some 85 % or 96% of the central mass of the cloud (Gaussian case), so that only a small proportion of pixels will be mapped on either of the endpoints of their class. This would make the K-S unimodality tests more discriminating as well.

Another point, meriting investigation by knowledgeable GIS data analysts, was mentioned in section 2.1. It concerns opportunities to minimize the risks of error propagation, when matching known

specimens of ground cover to the image by using supervised classification techniques, such as trained discriminant functions or neural networks. Instead of training and classifying straight away on the full image, one might apply first an unsupervised classification with UNSUP. Once the best class structure is determined that way, one could first match the known spectral data (*i.e.* spectral library, or spatially limited ground knowledge within the image) to the best fitting class or spatial region in the image and then generalize by matching only with pixels in the image belonging to that class. For instance, in crop estimation or oil spillage detection, it may be possible to identify a local spatial region in the image with that type of ground cover. The next step would then be to determine its (unsupervised) class membership and start generalizing over the image, using supervised matching techniques, restraining the matching only to those parts of the image which belong to that class.

Beyond such incremental improvements and data analytic validitions four main lines of methodological research beyond the present stage are suggested.

5.3.1. Innovative improvements in algorithmic performance and classification accuracy.

This focuses on methodological and algorithmic (precision, convergence and speed) enhancement of the basic features of unsupervised classification procedures. Accuracy evaluation and measurement is here associated with the errors of misclassification, known in standard supervised classification methods such as in parametric discriminant analysis and trained artificial neural networks. As these methods are based on, and presuppose the matching of image pixels to priorly known class or ground cover data, the concepts of misclassification and errors of misclassification can be immediately defined and modeled in, for instance, the well known *confusion matrices.*
However, in non-parametric, unsupervised classification these concepts do not have an immediate meaning because, by definition, no such prior knowledge exists as the whole procedure aims to detect, define and generate *a posteriori* the system of classes as contained in the spectral information of a particular image and its subsequent mapping to the pixels of that image.
Hence there is no direct analogue here to the familiar confusion matrix. Yet we think it may be possible to develop analogous measures and concepts at the two levels that matter in our system of unsupervised classification:

(1) the reliability of the detected class system (how stable are its detected classes under pixel sampling and picture window);

(2) the accuracy of the pixel class mapping (are similar pixels mapped to the same class?).

A different problem, of course, is how classes and pixel mappings in an image correspond with available, usually occasional ground knowledge. This concerns the validity of the obtained set of classes and its image mapping. This can only be assessed by the examination of classification results for images with substantive patches of quality ground knowledge and will serve to major purposes:

- calibration and tuning: do the known patches correspond to discernible classes, and which levels of classification parameters serve to get the best correspondence? (at what levels can clouds or vegetation be separated from water);
- generalization potential: does the coverage in a known patch correspond with that of other patches with the same class?

5.3.2. Spectral demixing within classes at the (sub-)pixel or class level.

In conventional classification methods, such as the ISODATA module in the ERDAS/IMAGINE system (Smith, Pyden and Cole, 1995), pixels are allocated to a class just in terms of a particular class label. In our system, in addition to that, classes are characterized in terms of a univariate distribution along a linear segment, as determined by the first principal component of their singular value decomposition. Pixels can thus be represented as particular distribution values within that class (the *blending* option). This makes it possible to attack mixed pixel blending as a mixture problem.
For instance, it was pointed out to us during our last meeting at CRREL RS/GISC that certain patches in the mountain area were labeled as belonging to the same class as water in the Painted Rock reservoir. Obviously, in actual applications one usually could seek to separate these classes by choosing a higher, more discriminating separation for class merging. This, however, would change the whole classification pattern and structure across the full image. Instead of that, in our case one could

restrict oneself just to the investigation of the distribution within that class looking for multimodality or other mixing components. Using statistical demixing techniques one could thus decompose ('demix') that single class into sub-classes corresponding to 'shadow' and regular water. This approach might also be feasible for the analysis of snow cover.

Another priority is the development of nonparametric models more general then the linear singular value decomposition within classes. We should investigate whether recent neural network theory based on radial basis functions (RBF's), and the use of spiking neurons can show the same promises here as were propagated elsewhere. Together with the use of evolutionary computation methods to search for models for demixing of clusters consisting of multiple classes, and the usage of Bayesian methods to exploit the spatial the spatial structure during pixel classification. Spatial structure is exploited by computing prior probabilities over a spatial neighborhood, and use these to compute posterior pixel classification probabilities.

5.3.3. Unsupervised change detection

Ultimately, we should meet the challenge of the application of these methods and techniques of unsupervised classification to the corresponding multi-image problem of detecting change in (disaster) areas, comparing two or more similar coordinated and registered images, such as, for instance, provided by time-sequential overpasses by LANDSAT TM RS imagery. Results in this respect might assist GIS-users in the area of emergency management.

6. References

Anselin, L. (1988). *Spatial econometrics: methods and models.* Dordrecht: Kluwer Academic.

Ball, G. B. (1965). Data analysis in the social sciences: what about the details? In: *Proceedings of the Fall Joint Computing Conference*, Washinton, DC: Spartan Books. 533-559.

Bellman, R. E. (1961). *Adaptive control processes.* Princeton, NJ: Princeton University Press.

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, **18**(9), 509-517.

Breiman, L., Friedman, J., Olshen, R., Stone, C. (1984). *Classification and Regression Trees.* London: Chapman and Hall.

Cheeseman, P. and Stutz, J. (1996). Bayesian Classification (AutoClass): Theory and Results. In: Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (eds). *Advances in knowledge discovery and data mining.* Cambridge, MA: AAAI Press / MIT Press.

Cheng, B. and Titterington, D. M. (1994). Neural networks: a review from a statistical perspective. *Statistical science,* **9**, 2-30.

Cherkassky, V., Friedman, J. H. and Wechsler, H. (eds). (1994). *From statistics to neural networks. Theory and pattern recognition applications.* Berlin: Springer-Verlag.

Cox, T. F. and Cox, M. A. A. (1994). *Multidimensional scaling.* London: Chapman & Hall.

Cressie, N. (1991). *Statistics for spatial data.* New York: Wiley.

Dasarathy, B. V. (ed.). (1991). *Nearest neighbor (NN) Norms: NN pattern classification techniques.* Los Alamos, CA: IEE Computer Society Press.

Daubechies, I. (1992). *Ten lectures on wavelets.* CBMS-NSF, Regional Conference Series in Applied Mathematics, 61, Philadelphia, PA: SIAM.

Dutilleuil, P. and Legendre, P. (1993). Spatial heterogeneity and heteroscedasticity: an ecologocal paradigm versus a statistical concept. *Oikos*, **66**, 152-171.

Eurostat, 1994. *New tools for spatial analysis.* Proceedings of the workshop, Lisbon, 18 to 20 November 1993. Luxembourg: Office for Official Publications of the European Communities.

Friedman, J. H., Bentley, J. L. and Finkel, R. A. (1977). An algorithm for finding best matches in logarithmic expected time. *ACM transactions on mathematical software*, **3**(3), 209-226.

Friedman, J. H., and Tukey, J. W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Comp.*, **23**, 881-889.

Hall, D. J. and Ball, G. B. (1965). *ISODATA: a novel method of data analysis and pattern classification.* Technical report. Menlo Park, CA: Stanford Research Institute.

Hall, D. J. and Khanna, D. (1977). The ISODATA method of computation for relative perception of similarities and differences in complex and real computers. In: Enslein, K., Ralston, A. and Wilf, H. S. *Statistical methods for digital computers.* New York: Wiley, 340-373.

Huber, P. J. (1985). Projection pursuit. *Annals of Statistics.* **13**, No. 2, 435-475.

Joliffe, I. T. (1986). *Principal component analysis.* Berlin: Springer-Verlag.

Jones, M. C., and Sibson, R. (1987). What is projection pursuit? *Journal of the Royal Statistical Society,* **Series A, 150**,*Part 1*, 1-36.

Kanellopoulos, I., Varfis, A., Wilkinson, G. G. and Megier, J. (1991). Classification of remotely-sensed satellite images using multi-layer perceptron networks. In: Kohonen, T., Mäkisara, K., Simula, O. and Vangas, J. (eds) (1991). *Artificial neural networks. Proceedings of ICANN-91, Vol 2.* Elsevier Science Publications. Amsterdam: North-Holland.

Kanellopoulos, I., Varfis, A., Wilkinson, G. G. and Megier, J. (1992). Land cover discrimination in SPOT HRV imagery using an artificial neural network - A 20 class experiment. *International journal of remote sensing*, **13,** 917-924.

Kaufman, L. and Rousseeuw, P. J. (1990). *Finding groups in data. An introduction to cluster analysis.* New York: Wiley.

Kemenade, C. H. M. van, La Poutré, J. A. and Mokken, R. J. (1997). Wavelet transforms in RS image analysis. Technical report. Amsterdam: University of Amsterdam (CCSOM) / CWI.

Kemenade, C. H. M. van, La Poutré, J. A. and Mokken, R. J. (1998a). Density-based unsupervised classification for remote sensing. Technical report. Amsterdam: University of Amsterdam (CCSOM) / CWI.

Kemenade, C. H. M. van, La Poutré, J. A. and Mokken, R. J. (1998b). UNSUP. A system for unsupervised classification of multispectral images in remote sensing. Technical report. Amsterdam: University of Amsterdam (CCSOM) /CWI.

Kemenade, C. H. M. van, La Poutré, J. A. and Mokken, R. J. (1999a). Density-based unsupervised classification for remote sensing. To appear in: Kanellopoulos, I., Wilkinson, G. G. and Moons. *Machine vision and advanced image processing in remote sensing*. Berlin: Springer-Verlag.

Kemenade, C. H. M. van, La Poutré, J. A. and Mokken, R. J. (1999b). Density-based unsupervised classification for multi-spectral imagery. To appear in: Stein, A. and Van der Meer, F. *Spatial statistics and remote sensing*. Dordrecht: Kluwer Academic Publishers.

Kohonen, T. (1997). *Self-organizing maps (2^{nd} ed.)*. Berlin: Springer-Verlag.

Koornwinder, T. H. (ed.). (1993). *Wavelets: An elementary treatment of theory and applications.* Singapore: World Scientific Publishing Co., Inc.

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm. *Ecology*, **74**, 1659-1673.

Loftsgaarden, D. O. and Quesenberry, C. P. (1965). A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics*, **36**, 1049-1051.

Meyer, I. (1990). *Ondelettes et operateurs I.* Paris: Hermann.

Mokken, R. J. (1996). Remote Sensing, Statistics and Artificial Neural Networks - Some R & D Perspectives. *Proceedings of the International Symposium on Spectral Sensing Research '95 (ISSSR), Theme: Crisis Support. 26 November - 1 December 1995, Melbourne, Victoria, Australia.* CD-ROM. Canberra: Australian Government Publishing Service.

Openshaw, S. (1994). What is GISable spatial analysis? In: Eurostat, (1994), *New tools for spatial analysis*. Proceedings of the workshop, Lisbon, 18 to 20 November 1993. Luxembourg: Office for Official Publications of the European Communities, 37-44.

Richards, J. A. (1993). *Remote sensing digital image analysis. An introduction. (2^{nd} ed.)*. Berlin: Springer-Verlag.

Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge: Cambridge University Press.

Silverman, B. W. (1986). *Density estimation.* London: Chapman and Hall.

Scott, D. W. (1992). *Multivariate density estimation.* New York: Wiley.

Smith, C., Pyden, N. and Cole, P., (1995). *Erdas field guide. 3rd edition.* Atlanta, GA: ERDAS, Inc.

Stutz, J., Cheeseman, P., Hanson, R. and Taylor, W. (1996). AutoClass - a Bayesian Approach to Classification. In: Skilling, J., and Sibisi, S. *Maximum Entropy and Bayesian Methods, Proceedings of the "14th International MaxEnt Workshop" MAXENT 94, Cambridge, United Kingdom, 1-5/8/94.* Dordrecht: Kluwer Academic Publishers.

Tanner, M. (1993). *Tools for statistical inference* (2nd Edition). Berlin: Springer- verlag.

Strang, G. (1989), Wavelets and dilation equations: a brief introduction, *SIAM Review*, **31** (1989), 614-627.

Tobler, W. (1979). Cellular geography. In Gale, S. and Olsson, G. (eds). *Philosophy in geography.* Dordrecht: Reidel, 379-386.

Weiss, S. M. and Kulikowsky, C. A. (1991). *Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning and expert systems.* San Mateo, CA: Morgan Kaufmann.

Annex to

Final Report (March 1998- October 1998)
Remote Sensing, Statistics and Artificial Neural Networks
contract no. N 68171 95 C 9124
contractor Prof.dr R.J. Mokken
ALL/CCSOM, PSCW, University of Amsterdam

1. Statement showing amount of unused funds at the end of the covered period

3rd Incrementally Funded Period September 97 - October 98	total	$	0.00
total unused funds at end of covered period		**$**	**0.00**

2. List of important property acquired with contract funds during this period

none